Power-Iterative Strategy for ℓ_p-ℓ_2 Optimization for Compressive Sensing: Towards Global Solution

Jie Yan and Wu-Sheng Lu
Department of Electrical and Computer Engineering
University of Victoria
Victoria, BC, Canada V8W 3P6
jyan@ece.uvic.ca, wslu@ece.uvic.ca

***Abstract*—We study nonconvex relaxation of the combinatorial ℓ_0-minimization for compressive sensing. In an ℓ_p-ℓ_2 minimization setting with $p < 1$, we propose an iterative algorithm with two distinct features: (i) use of a proximal-point (P-P) objective function composed of a convex quadratic term and an ℓ_p-norm term, and a fast parallel-based solver for global minimization of the P-P function in each iteration; and (ii) a power-iterative strategy that begins by solving a convex ℓ_1-ℓ_2 problem whose solution is then used to start next ℓ_p-ℓ_2 problem with p close to but less than one. The process continues with gradually reduced p until a target power p_t is reached. By simulations the algorithm is shown to offer considerable performance gain.**

I. Introduction

As an alternative and effective data acquisition paradigm, compressive sensing (CS) acquires a signal by collecting a relatively small number of linear measurements, and the signal is recovered with a nonlinear process [1], [2]. A central point in CS is to seek an accurate or approximate solution to an underdetermined linear system while requiring the solution to have fewest nonzero components. Regardless of whether or not the measurements are noise-free, the recovery problem can be solved in an ℓ_1-ℓ_2 formulation as

$$\min \ \lambda\|\mathbf{s}\|_1 + \|\mathbf{\Theta}\mathbf{s} - \mathbf{y}\|_2^2 \tag{1}$$

where $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$ with $\mathbf{\Phi}$ a measurement matrix and $\mathbf{\Psi}$ an orthogonal transform, and $\lambda > 0$ is a regularization parameter.

As a variant of the well-known basis pursuit (BP) [3], (1) is a nonsmooth, convex, unconstrained problem for which many efficient solution techniques exist [4]. The ℓ_1-ℓ_2 problem was traditionally treated using various classical iterative optimization algorithms [3], homotopy solvers [5], [6] and greedy techniques like matching pursuit and orthogonal matching pursuit [7]. However, these algorithms are often impractical in high-dimensional problems, as often encountered in image processing applications [4]. Over the past several years, a family of iterative-shrinkage algorithms have emerged as highly effective numerical methods for the above ℓ_1-ℓ_2 problem. Of particular interest is a proximal-point-function based algorithm known as the fast iterative shrinkage-thresholding algorithm (FISTA) developed in [8]. The FISTA is shown to provide a convergence rate of $O(1/k^2)$ compared to the rate of $O(1/k)$ by the well-known proximal-point algorithm known as the iterative shrinkage-thresholding algorithm (ISTA), while maintaining practically the same complexity as the ISTA. With slight increase in complexity, an enhanced version of FISTA with monotone convergence (known as MFISTA) was also proposed [9]. In addition, algorithms based on ℓ_p minimization with $0 < p < 1$ were proposed for improved reconstruction performance relative to that obtained by ℓ_1 minimization [10], [11].

In this paper, we study a nonconvex relaxation of the combinatorial ℓ_0-minimization for CS in an ℓ_p-ℓ_2 minimization setting, namely,

$$\min \ F(\mathbf{s}) = \lambda\|\mathbf{s}\|_p^p + \|\mathbf{\Theta}\mathbf{s} - \mathbf{y}\|_2^2 \tag{2}$$

with $0 < p < 1$. New algorithms for CS signal reconstruction based on ℓ_p-ℓ_2 optimization are proposed. The two new ingredients in the proposed algorithms that are responsible for the algorithm to achieve considerable performance gain are briefly described below.

(i) We associate problem (2) with an ℓ_p-ℓ_2 P-P objective function $Q_p(\mathbf{s}, \mathbf{b}_k)$ where $p < 1$, and minimize the nonconvex P-P function at each iteration. We are able to devise a fast formulation to secure the global minimizer of $Q_p(\mathbf{s}, \mathbf{b}_k)$. In particular, a parallel implementation is incorporated into the global solver and is shown to significantly accelerate the algorithm.

(ii) We develop an algorithm in the framework of MFISTA, called modified MFISTA (M-MFISTA), by replacing the conventional shrinkage solver for the ℓ_1-ℓ_2 P-P function with the global solver for $Q_p(\mathbf{s}, \mathbf{b}_k)$. Equipped with M-MFISTA, a power-iterative strategy is designed to reach a solution of (2) for a target power value p_t.

The proposed algorithms are evaluated by performing CS reconstruction of sparse signals with various values of p, and to compare with the well known basis pursuit (BP) algorithm [3] with $p = 1$.

II. Notation and Background

A. Signal acquisition and recovery with compressive sensing

Compressive sensing (CS) based signal acquisition computes M linear measurements of an unknown signal $\mathbf{x} \in R^N$ with $M < N$. This acquisition process can be described as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} \quad \text{with} \quad \mathbf{\Phi} = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \ldots \ \boldsymbol{\phi}_M]^T \tag{3}$$

where $\boldsymbol{\phi}_k \in R^N (k = 1, 2, \ldots, M)$. Suppose signal $\mathbf{x}$ is K-sparse with respect to an orthonormal basis $\{\boldsymbol{\psi}_j\}_{j=1}^N$ ($\boldsymbol{\psi}_j \in R^N$), then $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \boldsymbol{\Psi}\mathbf{s} \tag{4}$$

where $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1 \ \boldsymbol{\psi}_2 \ \ldots \ \boldsymbol{\psi}_N]$ is an orthogonal matrix and $\mathbf{s}$ is a K-sparse signal with $K \ll N$ nonzero elements. The CS theory mandates that if the matrix $\boldsymbol{\Theta} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ obeys the restricted isometry property (RIP) of order $2K$, i.e. the inequality

$$(1 - \delta_{2K})||\mathbf{s}||_2^2 \leq ||\boldsymbol{\Theta}\mathbf{s}||_2^2 \leq (1 + \delta_{2K})||\mathbf{s}||_2^2$$

holds for all $2K$-sparse vectors $\mathbf{x}$ with $\delta_{2K} < \sqrt{2} - 1$, then $\mathbf{s}$ can be exactly recovered via the convex optimization

$$\min \quad ||\mathbf{s}||_1 \tag{5a}$$

$$\text{subject to:} \quad \boldsymbol{\Theta}\mathbf{s} = \mathbf{y} \tag{5b}$$

and $\mathbf{x}$ is recovered by Eq. (4).

A sensing matrix $\boldsymbol{\Phi}$ obeys RIP of order $2K$ with $\delta_{2K} < \sqrt{2} - 1$ if it is constructed by (i) sampling i.i.d. entries from the normal distribution with zero mean and variance $1/M$, or (ii) sampling i.i.d. entries from a symmetric Bernoulli distribution (i.e. Prob$(\phi_{ij} = \pm 1/\sqrt{M}) = 1/2$), or (iii) sampling i.i.d. from other sub-Gaussian distribution, or (iv) sampling a random projection matrix $\mathbf{P}$ that is incoherent with matrix $\boldsymbol{\Psi}$ and normalizing it as $\boldsymbol{\Phi} = \sqrt{N/M}\mathbf{P}$, with $M \geq CK\log(N/K)$ and C a constant [12].

In practice, $\mathbf{x}$ is most likely only approximately K-sparse in $\boldsymbol{\Psi}$. In addition, measurement noise may be introduced in the sensing process as $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{e}$. A relaxed version of problem (5) permits a small deviation as

$$\min \quad ||\mathbf{s}||_1 \tag{6a}$$

$$\text{subject to:} \quad ||\boldsymbol{\Theta}\mathbf{s} - \mathbf{y}||_2 \leq \epsilon \tag{6b}$$

where ϵ stands for the permissible deviation. This problem was first discussed in [3] as basis pursuit (BP). In recent years, many applications in signal and image processing, such as denoising, inpainting, deblurring and compressive sensing all lead to a variant of problem (6) that mixes ℓ_1 and ℓ_2 expressions in the form of (1) where the constraint is replaced with a penalty term. The parameter λ replaces the threshold ϵ in (6), in governing the tradeoff between the reconstruction error and signal sparsity.

B. FISTA and MFISTA

We begin with reviewing an algorithm, known as the iterative shrinkage-thresholding algorithm (ISTA), which also bears the names of "proximal-point method" and "separable surrogate functionals method" [4]. A key step in its kth iteration is to approximate the objective function in (1) by an easy-to-deal-with upper-bound (up to a constant) convex function given by

$$Q_1(\mathbf{s}, \mathbf{s}_{k-1}) = \frac{L}{2}\|\mathbf{s} - \mathbf{c}_k\|_2^2 + \lambda\|\mathbf{s}\|_1 \tag{7}$$

where $\mathbf{c}_k = \mathbf{s}_{k-1} - \frac{1}{L}\nabla f(\mathbf{s}_{k-1})$, $f(\mathbf{s}) = \|\boldsymbol{\Theta}\mathbf{s} - \mathbf{y}\|_2^2$, and L is the smallest Lipschitz constant of $\nabla f(\mathbf{s})$, which is found to be $L = 2\lambda_{\max}(\boldsymbol{\Theta}\boldsymbol{\Theta}^T)$. The kth iteration of the ISTA finds the next iterate by minimizing $Q_1(\mathbf{s}, \mathbf{s}_{k-1})$. Because both terms in $Q_1(\mathbf{s}, \mathbf{s}_{k-1})$ are coordinate-separable, it can be readily verified that the minimizer of $Q_1(\mathbf{s}, \mathbf{s}_{k-1})$ can be calculated by a simple shrinkage of vector $\mathbf{c}_k$ with a constant threshold λ/L as

$$\mathbf{s}_k = \mathcal{T}_{\lambda/L}(\mathbf{c}_k)$$

where operator $\mathcal{T}$ applies to a vector pointwisely with $\mathcal{T}_a(c) = \text{sign}(c)\max\{|c| - a, 0\}$. Once iterate $\mathbf{s}_k$ is obtained, it is used to update vector from $\mathbf{c}_k$ to $\mathbf{c}_{k+1}$ and the shrinkage operator $\mathcal{T}$ is then applied to it to get the next iterate, and so on.

The ISTA iterations are of very low complexity, however, the algorithm only provides a slow convergence rate of $O(1/k)$. In [8] and [9], an algorithm known as the fast iterative shrinkage-thresholding algorithm (FISTA) was proposed and shown to provide a much improved convergence rate of $O(1/k^2)$ with the complexity of each iteration being practically the same as that of ISTA. The FISTA is built on ISTA with an extra step in each iteration that, with the help of a sequence of scaling factors t_k, creates an auxiliary iterate $\mathbf{b}_{k+1}$ by moving the current iterate $\mathbf{s}_k$ along the direction of $\mathbf{s}_k - \mathbf{s}_{k-1}$ so as to improve the subsequent iterate $\mathbf{s}_{k+1}$. The steps in the kth iteration of FISTA are outlined as follows with initial $\mathbf{b}_1 = \mathbf{s}_0$ and $t_1 = 1$.

1) Perform shrinkage

$$\mathbf{s}_k = \mathcal{T}_{\lambda/L}\{\frac{2}{L}\boldsymbol{\Theta}^T(\mathbf{y} - \boldsymbol{\Theta}\mathbf{b}_k) + \mathbf{b}_k\};$$

2) Compute $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$;

3) Update $\mathbf{b}_{k+1} = \mathbf{s}_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{s}_k - \mathbf{s}_{k-1})$.

Furthermore, by including an additional step to the FISTA iteration, the algorithm is enhanced to possess desirable monotone convergence [9]. The algorithm so modified is called the monotone FISTA (MFISTA).

III. Power-Iterative Algorithms for ℓ_p-ℓ_2 Optimization

We consider the ℓ_p-ℓ_2 optimization problem (2). The algorithm described below is developed within the framework of MFISTA. Unlike a typical ℓ_1-ℓ_2 proximal-point (P-P) objective function, we associate each iteration of the algorithm to a P-P objective function given by

$$\begin{aligned} Q_p(\mathbf{s}, \mathbf{b}_k) =& f(\mathbf{b}_k) + \langle \mathbf{s} - \mathbf{b}_k, \nabla f(\mathbf{b}_k)\rangle \\ &+ \frac{L}{2}\|\mathbf{s} - \mathbf{b}_k\|_2^2 + \lambda\|\mathbf{s}\|_p^p \end{aligned} \tag{8}$$

where $f(\mathbf{s}) = \|\boldsymbol{\Theta}\mathbf{s} - \mathbf{y}\|_2^2$. With $p < 1$, minimizing $Q_p(\mathbf{s}, \mathbf{b}_k)$ in (8) is a nonconvex problem. By taking the advantage of function $Q_p(\mathbf{s}, \mathbf{b}_k)$ being separable in its variable coordinates, we devise a fast solver to secure its global minimizer, and then incorporate the global solver into the framework of MFISTA.

Up to a constant, the problem of minimizing $Q_p(\mathbf{s}, \mathbf{b}_k)$ can be cast as

$$\min \hat{Q}_p(\mathbf{s}, \mathbf{b}_k) = \frac{L}{2}||\mathbf{s} - \mathbf{c}_k||_2^2 + \lambda||\mathbf{s}||_p^p \tag{9}$$

with $\mathbf{c}_k = \mathbf{b}_k - \frac{1}{L}\nabla f(\mathbf{b}_k)$. At a glance, the objective function in (9) differs from (7) only slightly with its ℓ_1 term replaced by an ℓ_p term. The ℓ_p variation is expected to improve the CS recovery performance because with $p < 1$ problem (9) provides a problem setting closer to the ℓ_0-norm problem. However, with $p < 1$ the problem in (9) becomes nonconvex, hence conventional technique like soft shrinkage fails to work in general. In what follows, we present an efficient parallel processing technique to find the global solution of (9).

A. Global solver for the ℓ_p-ℓ_2 problem (9)

The objective function in (9) consists of two terms, both of which are separable. Consequently, (9) is reduced to a series of N 1-D problems of the form

$$\min u(s) = \frac{L}{2}(s - c)^2 + \lambda|s|^p. \tag{10}$$

An algorithm for the global solution of (10) (hence (9)) is proposed in [13]. Based on this, a global solver of (10) for $c \geq 0$ can readily be generated. If we denote the solution of this solver by $z = \texttt{gsol}(c, L, \lambda, p)$, then it is evident that the global solution of (10) for $c < 0$ can be obtained as $z = -\texttt{gsol}(-c, L, \lambda, p)$. A drawback of this solution method is its low efficiency, especially for large-scale problems, as one needs to solve N 1-D problems. Below we describe an improved algorithm which employs a parallel processing technique to accelerate the global solver.

For description conciseness, denote by $\mathbf{a}.{*}\mathbf{b}$ the component-wise product of vectors $\mathbf{a}$ and $\mathbf{b}$; by $\mathbf{a}.^p$ the vector whose ith component is $|\mathbf{a}(i)|^p$; and by $\mathbf{1}$ and $\mathbf{0}$ the all-one and zero vectors, respectively. We use $[\mathbf{a} > \mathbf{b}]$ ($[\mathbf{a} < \mathbf{b}]$) to denote a vector whose ith component is 1 if $\mathbf{a}(i) > \mathbf{b}(i)$ ($\mathbf{a}(i) < \mathbf{b}(i)$) and 0 otherwise; and $[\mathbf{a} \geq \mathbf{b}]$ ($[\mathbf{a} \leq \mathbf{b}]$) is similarly defined. Let Λ be a length-K subset of $\{1, 2, ..., N\}$ and $\mathbf{b}$ be a vector of length K, we use $\mathbf{a}(\Lambda) = \mathbf{b}$ to denote a vector obtained by updating the components of $\mathbf{a}$, whose indices are in Λ, with those of $\mathbf{b}$; $\mathbf{c} \leftarrow \mathbf{c}(\Lambda)$ denotes a vector of length K that retains those components of $\mathbf{c}$ whose indices are in Λ. A step-by-step description of the algorithm is given in Table I where it is quite obvious that the data are processed in a vector-wise rather than component-wise manner. The parallel processing of data is made possible by taking the advantage of the separable structure of the objective function in (9) and playing a technical trick about the signs of $\mathbf{c}_k$ (see (9)) as illustrated for the scalar case (10) earlier.

We remark that the proposed ℓ_p-ℓ_2 solver is highly parallel with exception only in Step 3.4 where a total of $|\Omega|$ calls for 1-D solver `gsol` are made. Since $|\Omega|$ is typically much smaller than N, overall the complexity of the proposed solver is considerably reduced compared with that required by applying 1-D solver `gsol` N times.

Input Data	$\mathbf{c}_k$, L, λ and p.
Output Data	$\mathbf{z}_k = \text{argmin } \hat{Q}_p(\mathbf{s}, \mathbf{b}_k)$.
Step 1	Set $\boldsymbol{\theta} = \text{sign}(\mathbf{c_k})$ and $\mathbf{c} = \boldsymbol{\theta}.*\mathbf{c_k}$.
Step 2	If $p = 0$, compute $\boldsymbol{\vartheta} = \left[\frac{L}{2}\mathbf{c}.^2 > (\lambda \cdot \mathbf{1})\right]$, set $\mathbf{z} = \mathbf{c}.*\boldsymbol{\vartheta}$ and do Step 4; otherwise do Step 3.
Step 3	1. Compute $s_c = [\lambda p(1-p)/L]^{1/(2-p)}$, set $\boldsymbol{\vartheta} = [(s_c \cdot \mathbf{1}) < \mathbf{c}]$ and $\mathbf{z} = \boldsymbol{\vartheta}$. 2. Define $\Lambda = \{i : \boldsymbol{\vartheta}(i) = 1\}$ and update $\mathbf{c} \leftarrow \mathbf{c}(\Lambda)$. 3. Compute $\mathbf{v} = L(s_c \cdot \mathbf{1} - \mathbf{c}) + \lambda p s_c^{p-1} \cdot \mathbf{1}$, update $\boldsymbol{\vartheta} = [\mathbf{v} \geq \mathbf{0}]$ and set $\tilde{\mathbf{s}} = s_c \cdot \boldsymbol{\vartheta}$. 4. Define $\Omega = \{i : \boldsymbol{\vartheta}(i) = 0\}$. For each $i \in \Omega$, replace the ith component of $\tilde{\mathbf{s}}$ by the global solution of (10) over $[s_c, c]$ with $c = \mathbf{c}(i)$. 5. Set $\boldsymbol{\vartheta} = [\frac{L}{2}\mathbf{c}.^2 > \frac{L}{2}(\tilde{\mathbf{s}} - \mathbf{c}).^2 + \lambda\tilde{\mathbf{s}}.^p]$ and $\tilde{\mathbf{z}} = \tilde{\mathbf{s}}.*\boldsymbol{\vartheta}$. 6. update $\mathbf{z}(\Lambda) = \tilde{\mathbf{z}}$.
Step 4	$\mathbf{z}_k = \boldsymbol{\theta}.*\mathbf{z}$.

TABLE I
A PARALLEL ℓ_p-ℓ_2 SOLVER FOR GLOBAL SOLUTION OF (9)

B. M-MFISTA and a power-iterative strategy

By replacing the conventional shrinkage solver for the ℓ_1-ℓ_2 P-P function with the global ℓ_p-ℓ_2 solver presented in Sec. III-A in MFISTA, an algorithm, called modified MFISTA (M-MFISTA), for a (local) solution of problem (2) can be constructed. The algorithm is outlined as follows.

Input Data	λ, p, $\boldsymbol{\Theta}$ and $\mathbf{y}$.
Output Data	Local solution of problem (2).
Step 1	Take $L = 2\lambda_{\max}(\boldsymbol{\Theta}\boldsymbol{\Theta}^T)$ as the Lipschitz constant of ∇f. Set initial iterate $\mathbf{s}_0$ and the number of iterations K_m. Set $\mathbf{b}_1 = \mathbf{s}_0$, $k = 1$ and $t_1 = 1$.
Step 2	Compute minimizer $\mathbf{z}_k$ of (9) using the parallel global solver. Then update $t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2,$ $\mathbf{s}_k = \text{argmin } \{F(\mathbf{s}) : \mathbf{s} = \mathbf{z}_k, \mathbf{s}_{k-1}\},$ $\mathbf{b}_{k+1} = \mathbf{s}_k + (t_k/t_{k+1})(\mathbf{z}_k - \mathbf{s}_k) + [(t_k - 1)/t_{k+1}](\mathbf{s}_k - \mathbf{s}_{k-1}).$
Step 3	If $k = K_m$, stop and output $\mathbf{s}_k$ as the solution; otherwise set $k = k + 1$ and repeat from Step 2.

TABLE II
THE M-MFISTA

Although in each iteration the M-MFISTA minimizes the P-P function globally, a solution of problem (2) obtained by M-MFISTA is not guaranteed globally optimal because (2) is a nonconvex problem for $p < 1$. In what follows we propose a power-iterative strategy that promotes a local algorithm such as M-MFISTA to converge to a solution of (2), which is likely globally optimal.

The power-iterative strategy begins by solving the convex ℓ_1-ℓ_2 problem in (1) based on MFISTA [8], [9] where a conventional soft-shrinkage operation is carried out in each iteration. The global solution $\mathbf{s}^{(0)}$ is then used as the initial

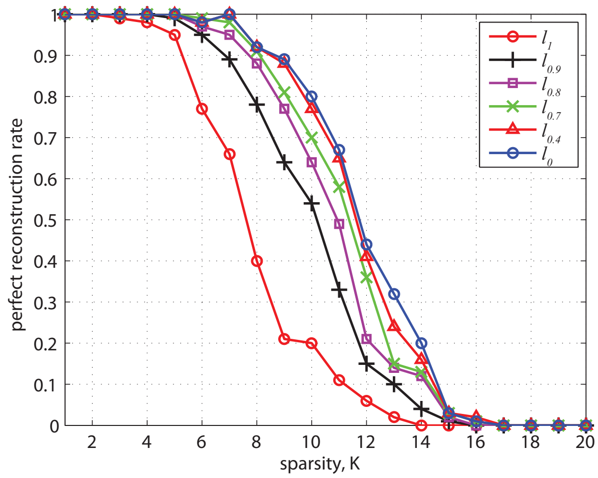

Fig. 1. Rate of perfect reconstruction for ℓ_p-ℓ_2 problems with $p = 1, 0.9, 0.8, 0.7, 0.4$ and 0 over 100 runs for signals of length $N = 32$ and number of random measurements $M = 20$.

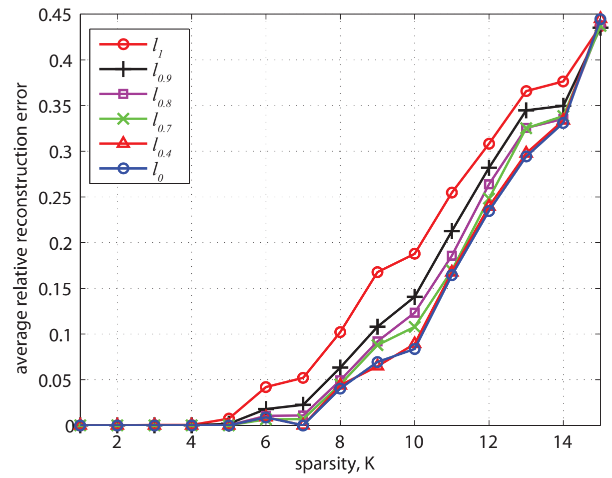

Fig. 2. Average relative reconstruction errors for ℓ_p-ℓ_2 problems with $p = 1, 0.9, 0.8, 0.7, 0.4$ and 0 over 100 runs for signals of length $N = 32$ and number of random measurements $M = 20$.

point to start the next ℓ_p-ℓ_2 problem with a p close to but slightly less than one. This problem is solved by the M-MFISTA and the solution obtained is denoted as $\mathbf{s}^{(1)}$. The iterate $\mathbf{s}^{(1)}$ is then served as an initial point for the next ℓ_p-ℓ_2 problem with p further reduced slightly. This process continues until the target power value p_t is reached.

For a nonconvex problem, a gradient based algorithm is not expected to converge to a global solution unless it starts at an initial point sufficiently close to the global solution. We argue that for a given power $p < 1$ and an appropriate value of λ, the global solution of (9) possesses continuity on p in the sense that the global solutions of (9) associated with powers p and $p + \Delta p$ are close to each other as long as the power difference Δp is sufficiently small in magnitude. It is based on this inuitive observation the above power iterative technique is developed to produce solutions of (2) that are likely globally optimal.

IV. Simulations

In this section, we evaluate the proposed algorithm for reconstruction of sparse signals by solving the ℓ_p-ℓ_2 problem with various power p.

Each K-sparse test signal $\mathbf{s}$ was constructed by assigning K values randomly drawn from $\mathcal{N}(0, 1)$ to K randomly selected locations of a zero vector of length $N = 32$. A total of 20 values of K from 1 to 20 were used. The number of measurements was set to $M = 20$ and a measurement matrix $\mathbf{\Phi}$ of size $M \times N$ was constructed with its elements drawn from $\mathcal{N}(0, 1)$ followed by normalizing each column to unit ℓ_2 norm. Matrix $\mathbf{\Psi}$ was set to the identity matrix as the test signals were all K-sparse. The power-iterative strategy in conjunction with M-MFISTA was applied to problem (2) to reconstruct $\mathbf{s}$. A sequence of power p was set from 1 to 0 with a decrement of $d = 0.1$. For each p, M-MFISTA was executed in a successive way with a set of decreasing λ's such that the equality constraint was practically satisfied, and a total of 50 M-MFISTA iterations was executed for each λ. A recovered signal $\hat{\mathbf{s}}$ was deemed perfect if the relative solution error $\|\hat{\mathbf{s}} - \mathbf{s}\|_2 / \|\mathbf{s}\|_2$ was less than 1e-5. For each value of K, the number of perfect reconstructions were counted over 100 runs.

Figs. 1 and 2 show the results for $p = 1, 0.9, 0.8, 0.7, 0.4$ and 0. It is observed that (i) for a fixed sparsity K, the rate of perfect reconstruction increases and the average relative reconstruction error reduces as a smaller power p was used. This justifies the usefulness of the proposed ℓ_p pursuit algorithm; (ii) the performance improvement tends to be nonlinear with respect to the change in power p, experiencing considerable improvement as p reduces from 1 to 0.9. As p decreases further, the performance continues to gain but the incremental gain becomes gradually less significant. It is also observed that the best reconstruction performance was achieved at $p = 0$.

Among other things, Figs. 3 and 4 compare the ℓ_0 (and $\ell_{0.9}$) solution obtained by the proposed method described above with an ℓ_0 (and $\ell_{0.9}$) solution obtained by M-MFISTA with the least-squares solution or the zero vector as the initial point, showing considerable performance gain achieved by the proposed method with an adequate initial point. This suggests that choosing an initial point greatly affects reconstruction performance. The simulations conducted so far seem to indicate that the proposed power-iterative method remains promising in approaching a global solution of the nonconvex problem (2).

V. Conclusions

A power-iterative strategy has been proposed for CS in an ℓ_p-ℓ_2 minimization setting. This methodology is built on a modified MFISTA (M-MFISTA) developed for local solution of the ℓ_p-ℓ_2 problem, in which a parallel global solver is devised for the ℓ_p-ℓ_2 P-P function. Experimental results for

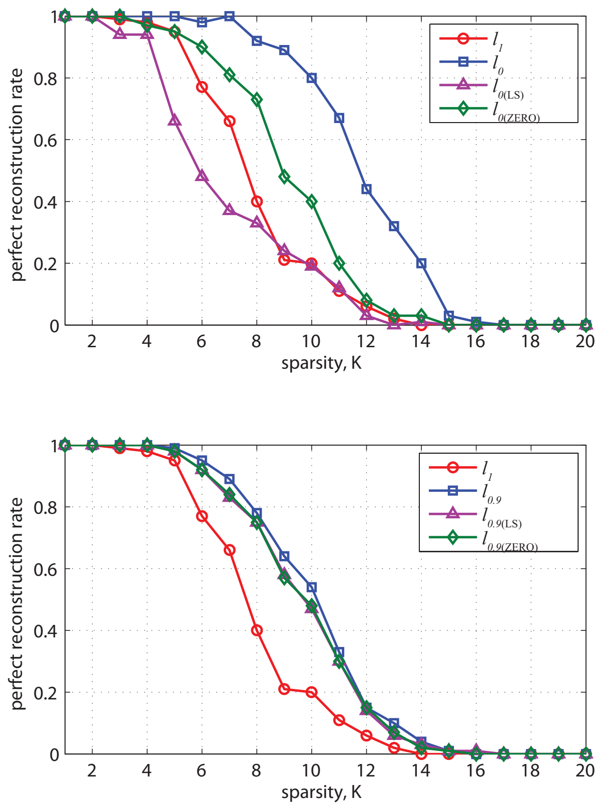

Fig. 3. Rate of perfect reconstruction for ℓ_p-ℓ_2 problems for $p = 0$ and 0.9 obtained with different initial points over 100 runs with $N = 32$ and $M = 20$. The upper graph compares the ℓ_0 solution obtained by the proposed method with the ℓ_0 solution obtained by M-MFISTA with the least-squares solution or the zero vector as the initial point. The lower graph does the comparison for the $p = 0.9$ counterpart. The curve corresponding to $p = 1$ is also shown as a comparison benchmark.

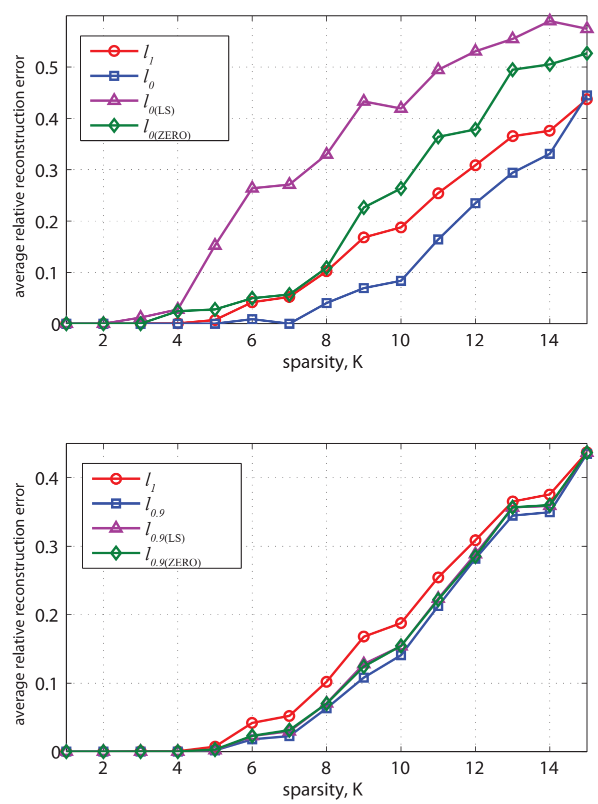

Fig. 4. Average relative reconstruction errors for ℓ_p-ℓ_2 problems for $p = 0$ and 0.9 obtained with different initial points over 100 runs with $N = 32$ and $M = 20$. The upper graph compares the ℓ_0 solution obtained by the proposed method with the ℓ_0 solution obtained by M-MFISTA with the least-squares solution or the zero vector as the initial point. The lower graph does the comparison for the $p = 0.9$ counterpart. The curve corresponding to $p = 1$ is also shown as a comparison benchmark.

CS signal recovery are presented to show the superiority of the proposed algorithms compared with the conventional BP benchmark, and to demonstrate that the solutions obtained are highly likely to be globally optimal.

ACKNOWLEDGEMENT

The authors are grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC) for supporting this work.

REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[4] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *Signal Processing Magazine, IEEE*, vol. 27, no. 3, pp. 76–88, 2010.

[5] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of statistics*, vol. 32, no. 2, pp. 407–451, 2004.

[6] D. Donoho and Y. Tsaig, "Fast solution of l1-norm minimization problems when the solution may be sparse," *IEEE Trans. Information Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.

[7] S. Zhang and S. Mallat, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[9] ——, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *Image Processing, IEEE Transactions on*, vol. 18, no. 11, pp. 2419–2434, 2009.

[10] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 707–710, 2007.

[11] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *ICASSP 2008*, pp. 3869–3872.

[12] E. Candès and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[13] J. Yan and W.-S. Lu, "New algorithms for sparse representation of discrete signals based on ℓ_p-ℓ_2 optimization," in *PacRim 2011*, submitted for publication.